

## Reference Linking nell'ambiente della biblioteca ibrida

### Parte 3: Generalizzazione della soluzione SFX nell'esperimento "SFX@Ghent & SFX@LANL"

Herbert Van de Sompel  
Los Alamos National Laboratory - Research Library  
[herbert.vandesompel@rug.ac.be](mailto:herbert.vandesompel@rug.ac.be)

Patrick Hochstenbach  
Automation Department of the Central Library University of Ghent, Belgium  
[hochsten@lanl.gov](mailto:hochsten@lanl.gov)

[Nota del Traduttore: si sono tralasciati Abstract e Acknowledgements, nonché le esemplificazioni della sezione "Illustrazione dei risultati del progetto", cfr. [testo originale](#); è stato aggiornato l'indirizzo e.mail di Hochstenbach]

## Riformulare il problema SFX

### Enunciazione del problema

E' opportuno che il problema SFX venga riformulato nel contesto dei convegni sul linking citazionale (*reference linking*), e delle pubblicazioni che ne sono seguite, organizzati dalla [Digital Library Federation](#) (DLF), la [National Information Standards Organization \(NISO\)](#), la [National Federation of Abstracting and Indexing Services \(NFAIS\)](#), e la [Society for Scholarly Publishing \(SSP\)](#) ([Caplan 1999a](#); [Caplan 1999b](#); [Caplan & Arms 1999](#); [Needleman 1999](#)).

L'enunciazione generica del problema del linking citazionale formulata dal Gruppo sul Reference Linking, è stata ([Caplan 1999a](#); [Caplan & Arms 1999](#)):

*"Data l'informazione di una citazione standard, come si arriva alla cosa a cui si riferisce?"*

Il Gruppo di lavoro si è però concentrato su una specifica variante di esso:

*"Data l'informazione di una citazione ad un articolo di periodico, come può l'utente arrivare dalla citazione ad una copia appropriata dell'articolo?"*

La ricerca SFX si è rivolta anch'essa a questi problemi, ma solo come ad un'istanza di un problema più generale:

*"Dati dei metadati bibliografici, come si possono presentare servizi estesi per essi?"*

### Metadati bibliografici come punto di partenza

Chiaramente, la ricerca SFX non si rivolge solo all'informazione in una citazione standard: il suo punto di partenza sono i metadati bibliografici in generale. In questa prospettiva, possono essere usati come punti di partenza in una specifica formulazione SFX, entità informazionali che si originano da tipiche risorse accademiche quali record di database bibliografici, Opac e archivi di preprint; vi rientrano anche le citazioni sia ad articoli di rivista sia a libri, trovate in articoli di periodici o libri; e in una specifica SFX sono un punto di partenza valido anche metadati parziali quali il nome di un autore preso da un messaggio e.mail.

### Servizi estesi come obiettivo

Una tale generalizzazione investe la tesi stessa del problema, in quanto la ricerca SFX non è interessata solo al collegamento al full-text relativo ad una data citazione in un articolo, ma mira alla presentazione, per ogni metadato utilizzato come punto di partenza, di una varietà di servizi estesi: per servizi estesi si intendono quelli che presentano un'entità informativa (definita link-

source) di una biblioteca digitale, nel contesto dell'intero ambiente informativo della biblioteca stessa ([Van de Sompel & Hochstenbach 1999a](#)).

Per esempio, per una data link-source che sia un record di un database bibliografico, i servizi estesi possono includere la presentazione di:

- il testo pieno dell'articolo indicizzato nella link-source
- un record di un altro database che contiene un abstract dello stesso articolo
- informazioni citazionali corrispondenti alla link-source (dati ISIS)
- il possesso della biblioteca per la rivista che contiene l'articolo della link-source.

### **Rilevanza globale e locale dei servizi estesi**

Nell'espressione "servizi estesi rilevanti" utilizzata nella enunciazione del problema SFX, è particolarmente importante l'aggettivo "rilevante": esso copre infatti due ambiti di significato: rilevanza come nozione globale e rilevanza come nozione locale. Per spiegarli, prendiamo in considerazione i seguenti tipi di servizi estesi:

- full-text: un servizio che fornisca il testo pieno a cui si riferisce la link-source
- review: un servizio che mostri la recensione al libro a cui si riferisce la link-source
- abstract: un servizio che fornisca un sunto, proveniente da un database di indicizzazione e abstract per una data link-source.

Rilevante come nozione globale può essere interpretato come il contrario di "irrilevante in ogni contesto": alcuni aspetti dei servizi estesi sono indipendenti dal contesto di una specifica collezione ma si applicano ad un livello globale:

- full-text: un servizio di full-text ha rilevanza globale se l'anno di pubblicazione di un articolo è posteriore a quello della prima uscita elettronica della rivista che lo contiene, mentre non ha senso in nessun caso presentare un servizio full-text per un articolo il cui anno di pubblicazione sia precedente all'anno della prima uscita elettronica della rivista che lo contiene. Quindi, l'anno di pubblicazione del primo fascicolo elettronico è un vincolo di rilevanza globale per il servizio full-text;
- review: è sempre irrilevante presentare il servizio di recensione a un libro se la link-source si riferisce ad un articolo di periodico; ma se la link-source descrive un libro, tale servizio di recensione è globalmente rilevante: per il servizio "recensione" dunque, è il tipo di materiale ad essere un vincolo di portata globale;
- abstract: il vincolo di portata globale che governa la rilevanza di un servizio abstract che cerchi all'interno di un particolare database secondario il riassunto per una data citazione, è che tale servizio è globalmente rilevante se il periodico che contiene l'articolo è indicizzato in quel database, altrimenti è globalmente irrilevante.

Rilevante come nozione locale si riferisce invece al fatto che ci sono altri aspetti dei servizi estesi che dipendono dall'ambito di una certa collezione digitale. La rilevanza locale ha due manifestazioni:

- rilevanza relativa al contenuto di una certa collezione locale: certi servizi, pur rilevanti a livello globale, diventano irrilevanti se la collezione della biblioteca digitale non include le risorse informative richieste per implementarli: anche se un servizio full-text è globalmente rilevante per una certa link-source, può essere considerato irrilevante nel contesto di una certa collezione digitale se la rivista a cui si riferisce la link-source non è parte della collezione; analogamente, sebbene un servizio abstract che punti ad un particolare database secondario possa essere di rilevanza globale, non avrà rilevanza locale se la biblioteca digitale dell'utente non fornisce accesso ad alcuna implementazione di tale database, mentre in caso contrario avrà rilevanza locale;

- rilevanza relativa all'implementazione della collezione locale: la rilevanza dei servizi estesi dipende anche dall'implementazione tecnica delle risorse informative coinvolte nei servizi: sebbene un servizio full-text sia globalmente rilevante (e cioè esista l'edizione elettronica di un articolo) e rilevante anche in relazione al contenuto di una certa collezione (gli utenti della biblioteca digitale sono autorizzati ad accedere all'edizione elettronica), potrebbe risultare inappropriato lasciare il link al servizio di full-text che punta al sito dell'editore quando la biblioteca ne ha una copia archiviata localmente: nella discussione del DLF sul reference linking, questo problema è stato definito "il problema di Harvard" ([Caplan 1999a](#)). Ma problemi analoghi si presentano anche nel più ampio ambito dei servizi estesi: per esempio, come sopra, un servizio abstract può essere globalmente rilevante (il periodico che contiene l'articolo è indicizzato in quel particolare database secondario), e rilevante anche in relazione al contenuto della collezione (la particolare biblioteca digitale fornisce accesso a quel database), tuttavia il servizio può essere irrilevante in relazione all'implementazione, se essa è tale che non supporta un meccanismo di link-to basato sui parametri necessari per la ricerca dell'abstract.

### Sistemi che supportano una risoluzione selettiva (ovvero la scelta di un risolutore locale)

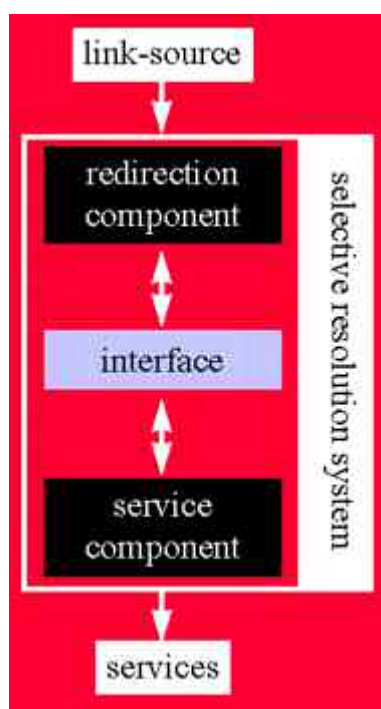


Figura 1: Sistemi che supportano la risoluzione selettiva

Entrambi gli aspetti relativi alla rilevanza locale dei servizi estesi indicano la necessità di soluzioni di linking aperte che, quando presentano i link all'utente, tengano conto della collezione locale ([Van de Sompel & Hochstenbach 1999a](#)). Anche la discussione all'interno del DLF sul reference linking, affrontando il problema di Harvard, si è riferita a soluzioni di linking aperte come ad un possibile supporto alla risoluzione selettiva ([Caplan & Arms 1999](#)). Risulta evidente come il problema della rilevanza locale dei servizi estesi non sia che una generalizzazione di certi aspetti del problema di Harvard: perciò, un approccio risolutivo che prenda in carico il problema più ampio conterrà elementi validi anche per affrontare il più limitato problema di Harvard.

In relazione al problema di Harvard, Caplan e Arms dividono i sistemi che supportano la risoluzione selettiva in due categorie:

- sistemi con un database di localizzazione (che cioè porta le configurazioni adatte all'ambiente locale) non locale, a cui le istituzioni forniscono un profilo che descrive le loro collezioni full-text: il profilo controlla la selezione di link che viene rimandata all'utente;

- sistemi con un database di localizzazione istituzionale, che descrive la collezione di testi pieni locale, e poi un database di localizzazione globale come riserva; in questo caso è presente anche un meccanismo per passare le richieste di risoluzione in prima istanza al risolutore locale e, solo nel caso di un full-text ivi non reperito, al risolutore globale.

Queste categorizzazioni possono essere ulteriormente generalizzate:

- allargando la portata dei servizi da fornire al di là del full-text e prendendo in considerazione tutti i servizi estesi;
- identificando i componenti software cruciali per i sistemi che supportano la risoluzione selettiva:
  - il meccanismo di reindirizzamento, che porta i metadati della link-source per cui si richiedono i servizi estesi, dalla risorsa informativa cui appartiene la link-source al componente del servizio: esso affronta il problema a cui ci si è riferiti come “cattura della link-source” ([Van de Sompel & Hochstenbach 1999a](#)).
  - il componente di servizio, che prende in input i metadati, provenienti da qualunque risorsa informativa nella collezione digitale, e fornisce in output i servizi estesi: esso è un’estensione del “database di localizzazione” cui si riferiscono Caplan e Arms;
- riconoscendo che l’ordine del reindirizzamento è variabile:
  - ci può essere prima il reindirizzamento dei metadati della link-source al componente locale di servizio, utilizzando poi un componente centrale di servizio per completare l’insieme dei servizi che possono essere offerti;
  - ci può essere il reindirizzamento dei metadati della link-source a un componente centrale di servizio, i cui servizi di default possono essere sovrascritti e/o completati in un secondo momento, dopo aver comunicato con il componente locale di servizio.

## CATEGORIA

Categoria 1		Centrale	Centrale
Categoria 2	a	Centrale & locale	locale => centrale
	b	centrale & locale	centrale => locale
Categoria 3		locale	locale
		<b>COMPONENTE DI SERVIZIO</b>	<b>ORDINE DI REINDIRIZZAMENTO</b>

Tavola 1: categorizzazione dei sistemi che supportano la risoluzione selettiva

La tavola 1 rappresenta la categorizzazione risultante e mostra tre categorie principali di sistemi che supportano la risoluzione selettiva, sulla base della natura del componente di servizio e sull’ordine di reindirizzamento:

- la categoria 1 ha solo un componente centrale di servizio, e perciò un meccanismo di reindirizzamento centrale. In un certo senso, questa è la categoria sotto cui rientra la soluzione [LinkOut](#) della NCBI, tuttavia, poiché tale soluzione è legata al database PubMed e non può essere usata in connessione con altre risorse, difficilmente si può parlare di vero e proprio componente di servizio nel senso sopra descritto;
- la categoria 2 ha sia un componente di servizio centrale che locale, che insieme contribuiscono alla presentazione dei servizi; c’è quindi anche una qualche forma di comunicazione tra i due. Per questa categoria si possono immaginare entrambi gli approcci di reindirizzamento menzionati sopra;
- la categoria 3 si costruisce unicamente su un componente locale di servizio e perciò necessita anche di un meccanismo di reindirizzamento locale. L’implementazione SFX sia

nell'esperimento Elektron che di Ghent ([SFX@Ghent](mailto:SFX@Ghent) & [SFX@LANL](mailto:SFX@LANL)) ricade in questa categoria.

## L'esperimento [SFX@Ghent](mailto:SFX@Ghent) & [SFX@LANL](mailto:SFX@LANL)

Nell'esperimento [SFX@Ghent](mailto:SFX@Ghent) & [SFX@LANL](mailto:SFX@LANL) (aprile-giugno 1999), d'ora in poi chiamato [Ghent&LANL](mailto:Ghent&LANL), il team [Library Without Walls](mailto:Library Without Walls) della Biblioteca di ricerca, al laboratorio nazionale di Los Alamos, ([LANL](mailto:LANL)) e il dipartimento automazione della biblioteca centralizzata dell'[Università di Ghent](mailto:Università di Ghent) hanno cooperato per verificare la fattibilità dell'approccio SFX come mezzo per fornire servizi estesi in un ambiente informativo complesso e realistico.

L'ambiente informativo in cui è stato condotto [Ghent&LANL](mailto:Ghent&LANL) è significativamente diverso da quello del primo esperimento di SFX, Elektron. Per chiarire, la tavola 2 presenta un quadro d'insieme delle risorse informative usate in [Ghent&LANL](mailto:Ghent&LANL): le righe mostrano i nomi delle risorse informative coinvolte nell'esperimento, le colonne si riferiscono alle collezioni digitali. Per ogni incrocio risorsa/collezione, la tavola indica:

- il tipo di risorsa (opac, database bibliografico, collezione di testi (FLTXT) o servizio web;
- l'autorità che gestisce la risorsa;
- se, all'interno della specifica biblioteca digitale, la risorsa è usata come punto di partenza dei link: se sì, ciò significa che le entità informative della risorsa possono essere link-source da cui richiedere dei servizi estesi, e quindi che l'autorità che gestisce la risorsa l'ha resa tale da supportare SFX;
- se, all'interno della biblioteca digitale, la risorsa è usata come target (punto di arrivo dei link): se sì, ciò significa che ad essa puntano i link che offrono servizi estesi, e quindi che l'autorità che gestisce la risorsa, per permetterle di essere il punto di arrivo dei link dinamici SFX, ha sviluppato una sintassi di link-to (di puntamento interno).

RISORSE	Tipo	GHENT			LANL		
		Authority	Source	Target	Authority	Source	Target
Advance	OPAC	-	-	-	LANL	yes	yes
Aleph 500	OPAC	Ghent	yes	yes	-	-	-
Amazon.com	WWW	Amazon	no	yes	Amazon	no	yes
Antilope	OPAC	UA	no	yes	-	-	-
APS PROLA	FTXT	APS	yes	yes	APS	yes	yes
the arXiv	FTXT	LANL	yes	yes	LANL	yes	yes
BIOSIS	A&I	Ghent	yes	no	LANL	yes	no
Books in Print	A&I	Ghent	yes	yes	Ghent	yes	yes
Compendex	A&I	Ghent	yes	no	LANL	yes	no
Current Contents	A&I	Ghent	yes	yes	Ghent	yes	yes
EconLit	A&I	Ghent	yes	no	-	-	-
Genome base	A&I	NCBI	no	yes	NCBI	no	yes
Inspec	A&I	-	-	-	LANL	yes	no
		SP	no	yes	SP	no	yes
Ulrich's	A&I	Ghent	yes	yes	-	-	-
LiSa	A&I	Ghent	yes	yes	-	-	-
MathSci	A&I	Ghent	yes	no	-	-	-
Medline	A&I	Ghent	yes	no	-	-	-
		NCBI	no	yes	NCBI	no	yes
SciSearch	A&I	LANL	yes	yes	LANL	yes	yes
ScienceServer	FTXT	LANL	no	yes	LANL	no	yes
Various	FTXT	various	no	yes	various	no	yes
Wiley InterScience	FTXT	Wiley	yes	yes	Wiley	yes	yes

Tavola 2: risorse informative nel progetto Ghent&LANL

Qualche considerazione sulla tavola 2:

- come si vede, alcune risorse sono disponibili in entrambe le collezioni digitali, ma implementate su piattaforme diverse: questo è il caso di BIOSIS e Compendex, che a Ghent girano sulla piattaforma ERL della SilverPlatter mentre a LANL (al momento dell'esperimento) usavano un'implementazione Geac, Advance;
- per gli scopi dell'esperimento, Ghent e LANL hanno condiviso alcune delle loro risorse: Ghent ha reso disponibile a LANL la sua versione dei Books in Print e dei Current Contents su ERL, mentre LANL ha dato accesso alla sua implementazione locale del Science Citation Index di ISI, e al suo ScienceServer che contiene le riviste di Elsevier a testo pieno.
- Ghent usa due versioni di Medline: una versione archiviata localmente come source (punto di partenza) e la versione pubblica NCBI come target (punto di arrivo). Similmente LANL usa due versioni di Geac: l'implementazione locale Geac Advance come source e l'implementazione ERL della SilverPlatter a Boston come target: le ragioni di queste peculiarità sono da ricercarsi nei limiti di tempo che hanno impedito lo sviluppo di una sintassi di "link-to" per le versioni locali.
- Di particolare rilievo è il fatto che alcuni periodici della collezione InterScience della Wiley, così come l'intero archivio PROLA dell'American Physical Society, sono stati resi SFX-aware: sia Ghent che LANL possono usare le citazioni presenti nel testo pieno di questi archivi come link-sources per i servizi SFX; inoltre, nel corso di questo esperimento, Wiley ha implementato una sintassi link-to che ha poi, nel corso del 1999, messo in produzione; per l'archivio PROLA tale sintassi link-to era già disponibile.
- Alcuni nomi di risorsa richiedono qualche spiegazione. Aleph 500 è il sistema integrato di gestione della biblioteca di Ghent, Advance è quello di LANL, mentre Antilope è il catalogo collettivo dei periodici del Belgio gestito dall'Università di Anversa. Il nome "Various" si riferisce ad una varietà di archivi di testi pieni a cui, in questo esperimento, sono possibili dei link. Questo è, tra gli altri, il caso dell'Academic Press, della Company of Biologists, di HighWire, Springer, American Chemical Society etc. l'ArXiv è l'implementazione locale del repository di e-print di Ginsparg, ArXiv, sviluppata dal team della "Library without walls" della biblioteca di LANL: anch'esso è stato reso SFX-aware.
- Come si vede da una lettura accurata della tavola 2, dal punto di vista di ognuna delle due biblioteche digitali, le risorse informative SFX-aware sono ampiamente distribuite: alcune risorse sono gestite dal centro di automazione della biblioteca, altre sono gestite "in remoto", attualmente da tre authorities esterne: dal punto di vista di Ghent tali authorities sono LANL, Wiley e American Physical Society; per LANL sono Ghent, Wiley e American Physical Society.

Da quanto sopra, si può concludere che questo esperimento è stato molto realistico, per il numero delle risorse coinvolte, la loro distribuzione ed il coinvolgimento di componenti di servizio SFX multipli.

## **La necessità di una generalizzazione dei componenti SFX**

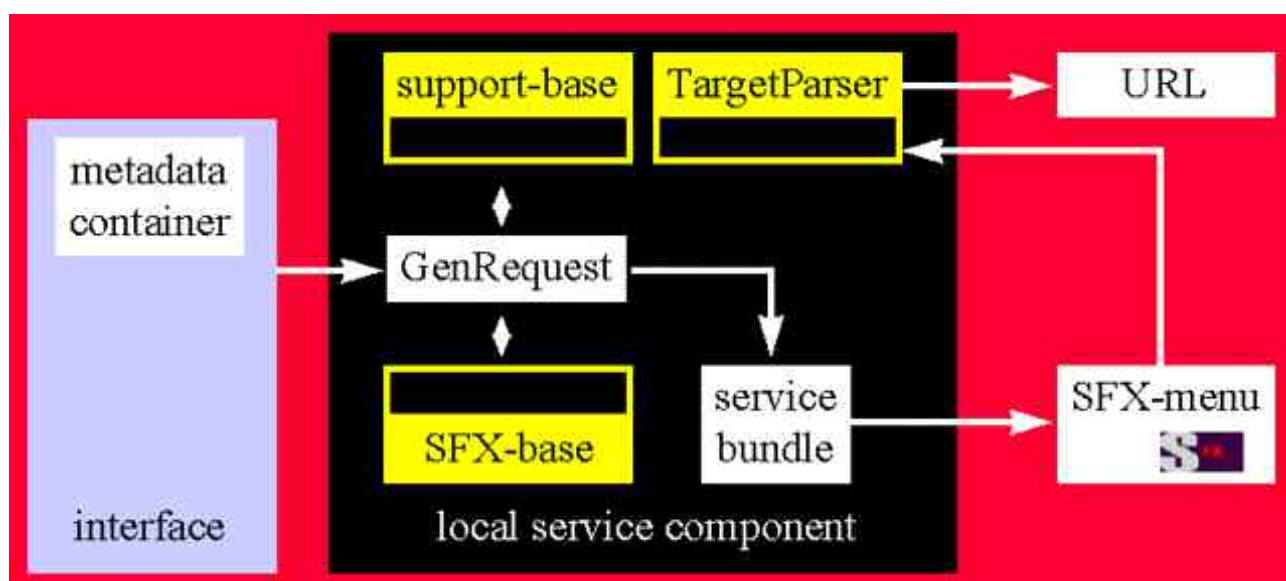
Sebbene i concetti fondamentali dell'SFX (linking dinamico, linking just-in-time e servizi concettuali) siano rimasti identici all'esperimento precedente (vedi [Van de Sompel & Hochstenbach 1999b](#)), la natura di questo ambiente di lavoro e i suoi scopi hanno condotto ad una forte generalizzazione dei componenti SFX. L'impulso principale che ha ispirato tale generalizzazione, e che distingue il progetto in oggetto da Elektron, sono:

- l'estensione della collezione digitale in cui è stato testato SFX, al di là della ristretta e controllata collezione della singola istituzione;



- l'estensione della portata dei dati per i quali si possono richiedere servizi estesi al di là delle collezioni gestite localmente: le link-sources, in questo esperimento, si originano anche da risorse gestite da authorities esterne;
- l'estensione del tipo di dati per cui possono essere richiesti i servizi estesi, al di là dei database bibliografici e degli Opac: le link-sources qui possono essere anche citazioni in un articolo;
- una strutturazione del link per servizi estesi verso risorse target basata sui metadati in generale e non solo sui metadati tipo SICI;
- la necessità di una alta portabilità della soluzione SFX tra le biblioteche digitali coinvolte.

La riprogettazione della soluzione SFX per Ghent&LANL ha condotto ad un'architettura con una chiara separazione tra il componente di reindirizzamento e il componente di servizio: entrambi i componenti ovviamente interoperano per ottenere un sistema funzionale, ma il componente di reindirizzamento può teoricamente operare in un ambiente con componenti di servizio non-SFX, mentre il componente di servizio SFX può funzionare ugualmente con un altro meccanismo di reindirizzamento, purché esso supporti l'invio dei metadati della link-source al componente di servizio SFX. Anche alcuni blocchi funzionali di entrambi i componenti sono stati generalizzati per affrontare i problemi originati dalla complessità dell'ambiente Ghent&LANL: l'approccio complessivo della soluzione generalizzata è illustrato nella figura 2 e sarà spiegato nei dettagli nel prosieguo di questo articolo. Le risorse informative che possono interoperare con SFX (d'ora in poi chiamate sistemi SFX-compatibili (SFX-aware)) inseriscono un bottone SFX per ogni link-source nelle risposte ad una query. Questo approccio just-in-time dell'SFX richiede all'utente di cliccare su quel bottone SFX quando vuole servizi estesi per una specifica link-source; in risposta a questo click, il componente locale di reindirizzamento SFX cattura i metadati della link-source (di solito) dalla risorsa origine, usando per fare ciò qualunque protocollo sia necessario. In seguito, i metadati della link-source, così come l'informazione sulla sua origine, saranno convertiti in un formato di interfacciamento. A questo punto, il sistema locale di reindirizzamento ha concluso il suo compito e può inviare questa informazione, in una rappresentazione coerente, al componente locale di servizio.



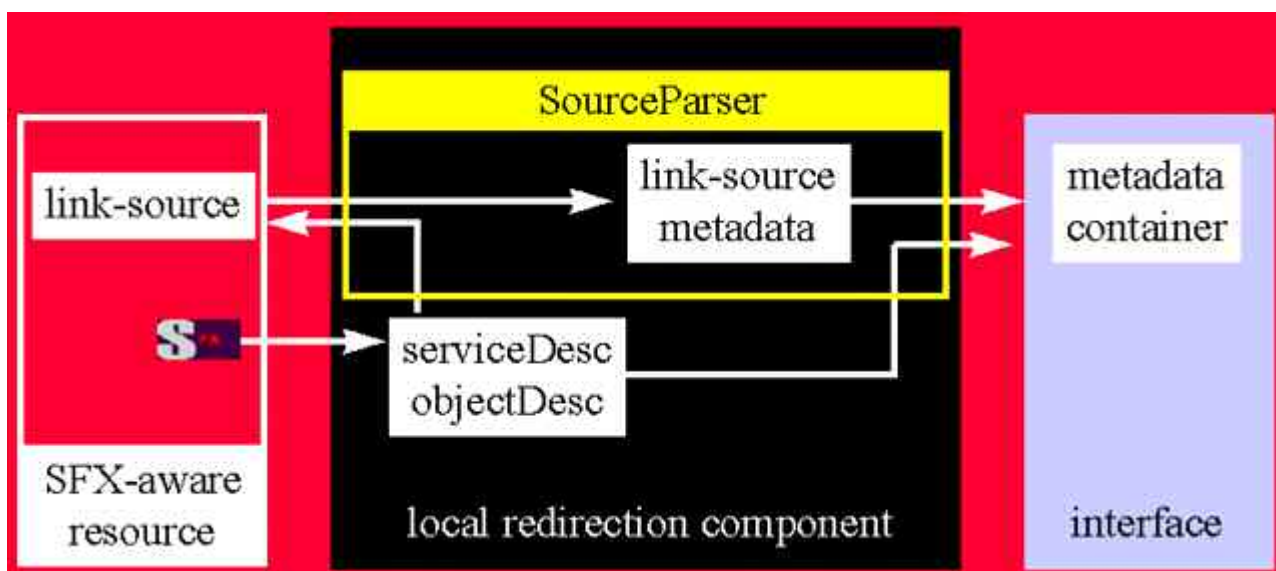


Figura 2: il componente locale di reindirizzamento e quello di servizio nella soluzione generalizzata SFX

Il primo compito del componente locale di servizio è analizzare/interpretare l'informazione passatagli dal componente locale di reindirizzamento, trasformandola in un oggetto di rappresentazione interna normalizzato. Durante questo processo, il contenuto originale può essere migliorato e/o aumentato. L'oggetto informativo risultante è quindi immesso nel processo valutativo SFX, in cui viene confrontato con il database SFX.

Il database SFX è un tipo particolare di database di linking: diversamente dai tradizionali servizi di linking, non contiene alcun link statico tra documenti (schede/citazioni/testi pieni); piuttosto, contiene una raccolta di servizi concettuali che esprimono le interrelazioni potenziali tra documenti, a livello della risorsa da cui originano. Il processo di valutazione SFX determina la rilevanza, per una data link-source, di ciascun servizio concettuale, usando il contenuto dell'oggetto informativo (o meglio la mancanza di determinati elementi). Poi, il risultante insieme di servizi rilevanti è reinviato all'utente nella schermata di menù SFX.

Coerentemente con l'approccio just-in-time di SFX, solo quando un utente sceglie un servizio di questo insieme, esso viene risolto nella URL a cui l'utente è reindirizzato.

## Il meccanismo SFX di indirizzamento locale

Il compito del meccanismo locale di reindirizzamento è trasportare i metadati della link-source al componente locale di reindirizzamento, che si interfaccia col componente locale di servizio. Per interoperare col meccanismo di reindirizzamento SFX, le authorities che gestiscono le risorse informative devono intervenire su di esse per renderle SFX-aware: ciò significa che la risorsa informativa va posta in grado di inserire, in ogni link-source del set di risposta ad una interrogazione, un bottone SFX che punti al componente locale di reindirizzamento.

Ecco alcune importanti considerazioni a questo riguardo, nel contesto di Ghen&LANL:

- molte risorse coinvolte nell'esperimento sono usate normalmente in produzione; questo significa che esse sono utilizzate anche da utenti che non hanno accesso ad un componente di servizio SFX; per evitare che a tali utenti sia presentato un bottone SFX irrilevante, una risorsa SFX-aware deve essere in grado di riconoscere se un utente ha accesso ad un componente di servizio SFX o no: sulla base di tale informazione, la risorsa inserisce o no il bottone SFX;
- alcune risorse informative sono accedute dagli utenti di entrambi gli ambienti informativi digitali: una risorsa SFX-aware deve essere in grado di far puntare il bottone SFX al componente locale di servizio appropriato, affinché esso possa inviare i metadati della link-source dalla risorsa informativa origine all'appropriato componente di servizio. Questo significa



che una risorsa SFX-aware deve essere in grado di parametrizzare il target del bottone SFX (cioè di passare come parametro la url a cui deve puntare il bottone SFX);

- c. alla richiesta di servizi estesi da parte dell'utente, il componente di reindirizzamento locale deve essere in grado di catturare i metadati della link-source dalla risorsa di origine. Questo significa che il componente di reindirizzamento locale deve essere informato dell'origine e dell'identità della link-source, per poter fare i passi opportuni. Data la quantità, distribuzione ed eterogeneità delle risorse SFX-aware nell'esperimento Ghent&LANL, si rende necessario un modo coerente di comunicare tali informazioni al componente di servizio;
- d. i metadati della link-source devono essere catturati da una gran varietà di risorse informative distribuite, che supportano diversi protocolli di accesso. Inoltre, tali risorse risponderanno inviando i metadati della link-source formattati secondo diversi schemi di metadati. Affinché il componente locale di reindirizzamento possa interfacciarsi in maniera generica (generalizzabile) con il componente locale di servizio, è desiderabile un unico formato di scambio dei metadati.

Come verrà mostrato nella descrizione dettagliata di seguito, questi problemi sono stati affrontati nei seguenti modi:

- per a. e b.: il meccanismo del CookiePusher
- Per c.: una struttura di URL coerente con SFX
- per d.: la soluzione del SourceParser (interprete della fonte)

## **Rendere una risorsa SFX-aware**

Le authorities che gestiscono le risorse informative devono migliorare i loro sistemi per renderli SFX-aware. La complessità dell'ambiente Ghent&LANL ha sollecitato una esplorazione ragionata dei possibili modi di rendere una risorsa SFX-aware, dato che può risultare accettabile e praticabile solo un approccio che minimizzi gli investimenti necessari da parte delle authorities.

Nella corrente implementazione del meccanismo di reindirizzamento SFX, le authorities devono:

- installare lo script Cookies-Pusher fornito dai manager di progetto Ghent&LANL
- rendere ipertestuale il bottone SFX nella link-source, usando una URL che rispetta un formato stabilito.

## **Il CookiePusher**

Lo script CookiePusher è una soluzione pragmatica introdotta per informare dinamicamente una risorsa informativa dell'esistenza e localizzazione, nell'ambiente dell'utente che la consulta, di un componente locale di reindirizzamento SFX. L'idea sottostante è che, se la URL del componente di reindirizzamento locale è scritto come cookie nel browser dell'utente che consulta una risorsa informativa, tale risorsa può in ogni momento accedervi; la disponibilità di questa URL è essenziale, dato che la risorsa deve essere in grado di indirizzare dinamicamente il bottone SFX all'appropriato componente locale, e tuttavia, per ragioni di sicurezza e privacy, i cookies sul browser possono essere letti solo da istanze provenienti dal dominio internet del server che ha mandato il cookie (vedi [Shishir 1996](#) pp. 203-204), perciò è impossibile settare tali cookies in modo tale che possano essere letti da tutti i sistemi informativi di una biblioteca digitale, dal momento che essa comprende risorse distribuite in diversi domini, tipicamente sia locali che remote rispetto all'istituzione dell'utente.

Per risolvere questo problema, bisogna che il primo passo, nella connessione, sia chiedere ad un server del dominio della risorsa informativa di creare un cookie HTTP: questa deviazione è chiamata CookiePusher. Il semplicissimo script CookiePusher è inserito nel dominio della risorsa informativa che deve essere resa SFX-aware: l'utente (attraverso la homepage approntata dalla sua istituzione), invece di connettersi direttamente alla URL della risorsa informativa, fa prima una

connessione al CookiePusher della risorsa, inviandogli i due parametri richiesti dallo script CookiePusher:

- localizzazione SFX: l'URL del componente locale di reindirizzamento
- reindirizzo: l'URL della risorsa desiderata.

Al ricevimento di questi parametri, il CookiePusher, prima leggerà la URL del componente di reindirizzamento locale e la userà per mettere un Cookie nel browser dell'utente, e poiché il cookiePusher è nel dominio della risorsa, quel cookie sarà leggibile dalla risorsa; poi il CookiePusher reindirizzerà l'utente alla URL della risorsa desiderata.

Quindi, una volta che il CookiePusher è stato installato, la URL per connettersi a quella risorsa deve essere:

**CookiePusher\_URL?SFX\_location= local\_SFX& Redirect= service\_URL**

dove:

- **CookiePusher\_URL** è la URL dello script CookiePusher
- **local\_SFX** è la URL del componente di reindirizzamento locale SFX
- **ervice\_URL** è la URL della risorsa informativa desiderata, quella che si usa normalmente in condizioni non SFX; tale URL può puntare all'iniziale schermo di ricerca di un database bibliografico, può essere la URL di un articolo sul sito di un editore, etc.

Sia **local\_SFX** che **service\_URL** sono codificate come URL. Per esempio:

[http://publish.aps.org/edaccess/prolatest/cookiepusher? SFX\\_location=http%3A%2F%2Fsiserv.rug.ac.be%2Fcgi-bin%2Fsfx%2Fbin%2Fmenu.cgi  
&Redirect=http%3A%2F%2Fpublish.aps.org%2Fedaccess%2Fprolatest%2Ftext%2FPRD%2Fv52%2Fi1%2Fp15\\_1](http://publish.aps.org/edaccess/prolatest/cookiepusher?SFX_location=http%3A%2F%2Fsiserv.rug.ac.be%2Fcgi-bin%2Fsfx%2Fbin%2Fmenu.cgi&Redirect=http%3A%2F%2Fpublish.aps.org%2Fedaccess%2Fprolatest%2Ftext%2FPRD%2Fv52%2Fi1%2Fp15_1)

è la URL usata per connettersi ad un item nel dominio APS/PROLA: il loro CookiePusher leggerà la localizzazione del componente di reindirizzamento locale dal parametro *SFX\_location* e lo utilizzerà per mettere un cookie chiamato *Local\_SFX* con valore:

<http%3A%2F%2Fsiserv.rug.ac.be%2Fcgi-bin%2Fsfx%2Fbin%2Fmenu.cgi>

che è la localizzazione codificata del componente di reindirizzamento locale di Ghent. Poi reindirizzerà l'utente alla localizzazione desiderata dall'utente in APS/PROLA:

[http://publish.aps.org/edaccess/prolatest/text/PRD/v52/i1/p15\\_1](http://publish.aps.org/edaccess/prolatest/text/PRD/v52/i1/p15_1)

Da ora in poi, in ogni momento della consultazione, APS/PROLA sarà in grado di leggere questo cookie e di puntare, se del caso, al componente di reindirizzamento di Ghent.

### Una struttura di URL coerente con SFX

L'essenziale del giro fatto con il CookiePusher è la capacità che esso conferisce ad una risorsa informativa di sapere in ogni momento se l'utente che la sta consultando ha accesso a un sistema di risoluzione selettivo, e se sì, qual è l'indirizzo del suo componente di reindirizzamento. Sulla base di tale informazione, la risorsa può decidere dinamicamente se inserire un bottone SFX o no nei risultati della ricerca, e se lo fa, a quale componente di reindirizzamento far puntare il bottone SFX. Per rendere i diversi sistemi coinvolti nell'esperimento interoperabili con SFX, alle authorities che gestiscono i sistemi è stato chiesto di rendere la URL a cui punta il bottone SFX (la URL SFX), conforme al seguente formato:

generale:	target?serviceDesc&objectDesc
dettagliata:	local_SFX?vendorID=<il fornitore>&databaseID=<la base>&objectDesc=<l'identificatore>

**Tavola 3: la sintassi della URL SFX**

dove:


- *target* è l'URL del componente locale di reindirizzamento della soluzione SFX
- *serviceDesc* definisce in maniera univoca la risorsa origine; contiene informazioni sul fornitore della risorsa e sulla risorsa stessa; questa informazione giocherà un ruolo cruciale nelle successive fasi del meccanismo locale di reindirizzamento SFX, così come nella base SFX, che è centrale al componente di servizio SFX.
- *objectDesc* contiene informazioni relative all'identità della link-source; la sua sintassi e il suo contenuto sono estremamente flessibili e saranno definite dall'authority che gestisce la risorsa,

in dipendenza dal fornitore e dalla implementazione del database; l'*objectDesc* tipicamente contiene l'identificatore univoco di record che identifica la link-source nella sua risorsa di origine; in alternativa, o in aggiunta, può contenere metadati tipo SICI; in certi casi, può contenere tutti i metadati della link-source.

- i valori dei parametri <il fornitore>, <la base>, <l'identificatore>, sono codificati come URL.

Le figure 3 e 6 mostrano esempi di link-sources provenienti dalle collezioni di Ghent o di LANL, e mostrano la loro URL SFX. Per ragioni di leggibilità, i valori dei parametri non sono mostrati nella loro codifica come URL, ma è piuttosto evidenziato quali parti dovrebbero essere codificate come URL con la funzione (l'etichetta) URLEncode.

#### □ Record 2 of 33 in Biological Abstracts 1999/01-1999/03



TI Long term outcome of patients with **hairy cell leukemia** treated with pentostatin.  
 AU [Ribeiro-Patricia](#); [Bouaffia-Fadhela](#); [Peaud-Pierre-Yves](#); [Blanc-Michel](#); [Salles-Bruno](#)  
 AD {a} Serv. Hematol., Cent. Hosp. Lyon-Sud, 165 Chemin du Grand Revoyet, 69495  
 SO [Cancer](#)-Jan. 1, 1999, 85 (1) 65-71..  
 PY 1999  
 DT Article-  
 IS 0008-543X  
 LA English  
 AI Y  
 ST Hominidae-, Primates-, Mammalia-, Vertebrata-, Chordata-, Animalia-  
 RN 53910-25-1: PENTOSTATIN  
 AN 199900063465  
 UD 19990223  
 AV 

La URL SFX per questa link-source, che punta al componente di reindirizzamento locale di Ghent, è:

[http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorId=ERL&databaseId=BX  
&objectDesc=URLEncode\(BX02 A:199900063465 I:0008-543X V:00085 S:000001 P:000065 Y:1999\)](http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorId=ERL&databaseId=BX&objectDesc=URLEncode(BX02 A:199900063465 I:0008-543X V:00085 S:000001 P:000065 Y:1999))

Nella parte serviceDesc della URL, *ERL* si riferisce all'implementazione ERL della SilverPlatter di BIOSIS, mentre *BX* è il nome dei databases BIOSIS in ambiente ERL. Il componente objectDesc contiene vari elementi informativi in una rappresentazione etichettata e a lunghezza fissa. BX02 è il volume del database BIOSIS da cui si origina la link-source, mentre 199900063465 è il numero di accesso cioè un numero di record unico per la link-source in BIOSIS. Altri elementi del objectDesc sono numero ISSN, volume, fascicolo, pagina iniziale e anno di pubblicazione.

**Figura 3: una link-source dall' implementazione ERL di BIOSIS a Ghent, e la sua SFX-URL**

Record 3 of 91		Mark	Full Record
Article:	 <a href="#">Article</a>		
Title:	<a href="#">The identification of cDNAs that affect the <b>mitosis</b>-to-<b>interphase</b> transition in <i>Schizosaccharomyces pombe</i>, including <i>sbp1</i>, which encodes a <i>spilp</i>-GTP-binding protein.</a>		
Author:	<a href="#">HE, XIANGWEI</a> ; <a href="#">HAYASHI, NAOYUKI</a> ; <a href="#">WALCOTT, NATHAN G.</a> ; <a href="#">AZUMA, YOSHIO</a> ; <a href="#">PATTERSON, THOMAS E.</a> ; <a href="#">BISCHOFF, F. RALF</a> ; <a href="#">NISHIMOTO, TAKEHARU</a> ; <a href="#">SHELLEY, SHELLEY</a>		
Source:	<a href="#">Genetics, Feb. 1998, v.148, no.2, p.645-656.</a>		
Other Links:			
<hr/>			
Location	Holdings		
MAIN	Holdings: v.109- (1985- ) ; Missing: v. 134 no. 4 (1993) ; Last rec'd: VOL.152 NO.2 / J issue on display ; Shelved as: GENETICS.		
WWW	<a href="http://www.genetics.org/">http://www.genetics.org/</a> ; Holdings: v.148, no.1- (Jan.1998- ) ; Abstracts and table of contents: v.148, no.2-v.147, no.4 (Feb.1980-Dec.1997)		

La URL SFX per questa link-source, che punta al componente di reindirizzamento di LANL, è :  
[http://vole.lanl.gov/cgi-bin/sfx/bin/menu.cgi?vendorId=ADVANCE&databaseId=Biosis&objectDesc=URLencode\(fetchId=21179970&objectId=PREV199800135979&SICI=0016-6731\(1998\)148:2<645:TIOCTA>2.0.TX;2-P\)](http://vole.lanl.gov/cgi-bin/sfx/bin/menu.cgi?vendorId=ADVANCE&databaseId=Biosis&objectDesc=URLencode(fetchId=21179970&objectId=PREV199800135979&SICI=0016-6731(1998)148:2<645:TIOCTA>2.0.TX;2-P))

La parte serviceDesc di questa URL è autoevidente. Il componente objectDesc è etichettato e i campi possono avere lunghezza variabile. Il fetchId è il numero unico della link-source nella implementazione di BIOSIS a LANL, mentre la parte objectId dopo "PREV" è il numero di accesso di BIOSIS che corrisponde al campo A nell'objectDesc della SilverPlatter in [Figure 3](#). La parte SICI contiene un SICI per la link-source, da cui è possibile derivare ISSN, volume, fascicolo, paginazione e anno di pubblicazione.

Figura 4: una link-source dall'implementazione di BIOSIS a LANL, Advance, e la sua SFX-URL



WebSPIRS: Search - Netscape

Special Effects: University of Ghent - Netscape

Article Abstract - Netscape

WILEY  
InterScience®

PERSONAL HOMEPAGE JOURNAL FINDER SEARCH HELP CONTACT US LOGOUT

ALL JOURNALS PREVIOUS ARTICLE NEXT ARTICLE

Article Abstract

**CANCER**

Online ISSN: 1097-0142 Print ISSN: 0008-543X

**Cancer**  
Volume 85, Issue 1, 1999. Pages: 65-71

ADD HOT ARTICLE

Original Article

Long term outcome of patients with hairy cell leukemia treated with pentostatin

Patricia Ribeiro, M.D.<sup>1</sup>, Fadhela Bouaffia, M.D.<sup>1</sup>, Pierre-Yves Peaud, M.D.<sup>2</sup>, Michel Blanc

References

- 1 Saven A, Piro L. Treatment of hairy cell leukemia. *Blood* 1992; **79**: 1111-20. [Medline](#) 
- 2 Jaiyesimi I, Kantarjian H, Estey E. Advances in therapy for hairy cell leukemia. A review. *Cancer* 1993; **72**: 5-16. [Medline](#) 
- 3 Saven A, Piro L. The newer purine analogues for the treatment of hairy-cell leukemia. *N Engl J Med* 1994; **330**: 691-7. [Medline](#) 

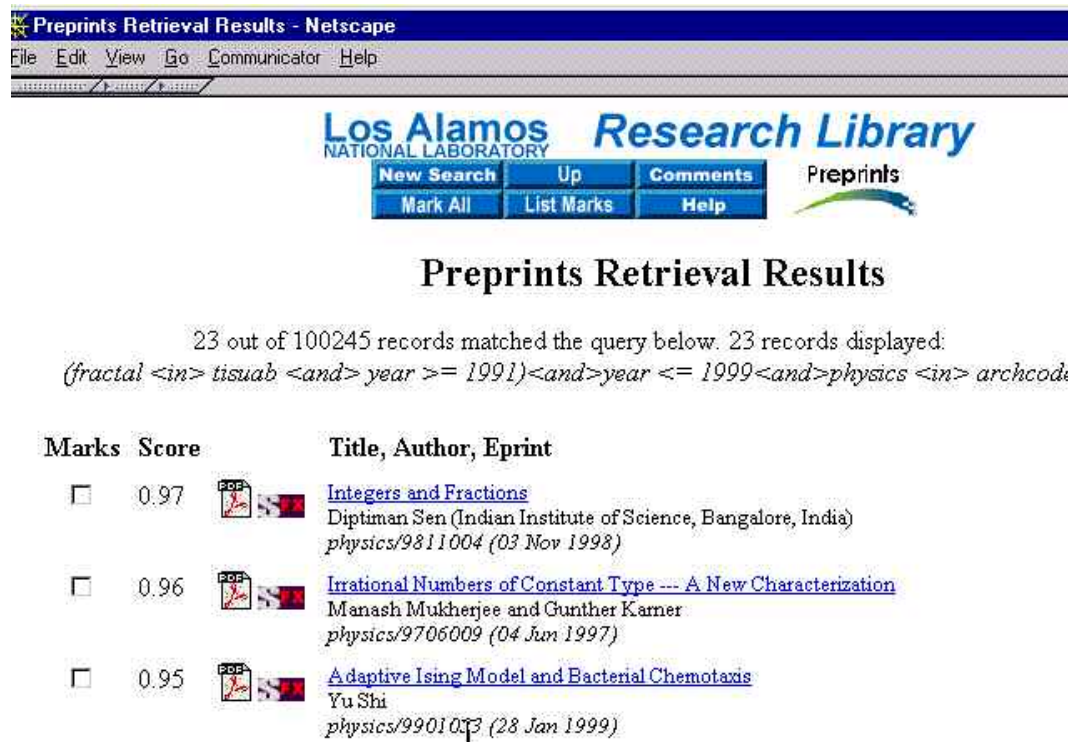


La URL SFX per la terza citazione presa come link-source, che punta al componente locale di reindirizzamento di Ghent, è:

[http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorId=Wiley&databaseId=WIS&objectDesc=URLencode\(TYPE=JCIT&SNM=Saven&FNM=A&SNM=Piro&FNM=L&ATL= The newer purine analogues for the treatment of hairy-cell leukemia.&JTL=N Engl J Med &PYR=1994&VID=330&PPF=691&PPL=7\)](http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorId=Wiley&databaseId=WIS&objectDesc=URLencode(TYPE=JCIT&SNM=Saven&FNM=A&SNM=Piro&FNM=L&ATL=The%20newer%20purine%20analogues%20for%20the%20treatment%20of%20hairy-cell%20leukemia.&JTL=N%20Engl%20J%20Med%20&PYR=1994&VID=330&PPF=691&PPL=7))

Il componente serviceDesc ora si riferisce alla collezione Wiley InterScience. L'objectDesc è etichettato e comincia con un'indicazione del tipo di materiale della citazione -- in questo caso citazione da rivista -- seguito da una ripetizione etichettata dell'intera citazione.

Figura 5: una link-source da Wiley InterScience e la sua SFX-URL



La URL SFX per la prima link-source nella schermata di risultati sopra riportati, che punta al componente locale di reindirizzamento di LANL, è:

[http://vole.lanl.gov/cgi-bin/sfx/bin/menu.cgi?vendorId=LANLTopic&databaseId=arXiv&objectDesc=URLencode\(fetchId=phys-9811004&objectId=physics/9811004\)](http://vole.lanl.gov/cgi-bin/sfx/bin/menu.cgi?vendorId=LANLTopic&databaseId=arXiv&objectDesc=URLencode(fetchId=phys-9811004&objectId=physics/9811004))

Il serviceDesc si riferisce alla implementazione locale di LANL dell'archivio di e-print di Ginsparg. Il fetchId è la chiave unica del record in tale implementazione, mentre l'objectId, molto simile, è il numero unico di record nell'implementazione di Ginsparg. Non c'è nessun altro metadato disponibile nel objectDesc.

Figura 6: una link-source da arXiv e la sua SFX-URL

## Catturare i metadati della link-source da una risorsa informativa SFX-aware con un SourceParser

Il meccanismo CookiePusher mette in grado una risorsa di inserire un bottone SFX per ogni link-source che viene trasferita ad un utente che la consulta. La struttura della URL SXF che viene puntata da questi bottoni SFX è stata resa omogenea fra le varie risorse nella forma: target?serviceDesc&objectDesc.

Quando un utente richiede i servizi estesi cliccando tale bottone SFX, viene inviata una richiesta al suo componente locale di reindirizzamento SFX, che riceverà i valori serviceDesc e objectDesc come parametri per il suo script così attivato; il componente locale mantiene una raccolta di script SourceParser (per l'analisi della link-source) con nomi corrispondenti ai serviceDesc validi (vedi tavola 4): dopo aver analizzato l'informazione del serviceDesc, lo script lancerà l'appropriata

procedura SourceParser. Questa procedura, specifica per ogni *serviceDesc*, implementa univocamente:

- l'interpretazione dell'informazione contenuta nel parametro *objectDesc*, sulla base della sintassi definita dal fornitore (vedi figure da 3 a 6);
- il meccanismo per catturare la link-source dalla sua risorsa di origine, sulla base del parametro "*origin*" e del contenuto dell'*objectDesc*. La tavola 4 mostra questi meccanismi di cattura per gli esempi delle figure da 3 a 6: come si vede, nessuna cattura effettiva è richiesta per le citazioni Wiley poiché queste sono completamente trasferite nella parte *objectDesc* della URL SFX, e la stessa tecnica è usata per le citazioni dall'archivio PROLA. Diversamente, le implementazioni BIOSIS di Ghent e di LANL inviano entrambe alcuni metadati (correlati al SICI) nel *objectDesc*, ma poiché alcuni servizi estesi che SFX mira a fornire richiedono metadati aggiuntivi, ci vuole una cattura per ottenere informazioni più complete. Per arXiv infine, poiché l'*objectDesc* contiene solo un identificatore, ci vuole assolutamente una cattura.
- la conversione dei metadati della link-source, che sono espressi nello schema di metadati supportato dalla authority che gestisce la risorsa di origine, verso uno schema/contenitore di metadati conforme al formato univoco di scambio; questo schema di metadati è l'interfaccia tra il componente locale di reindirizzamento e il componente locale di servizio.

RESOURCE	ServiceDesc		SourceParser	Fetch protocol	Fetch key
the arXiv	LANLTopic	arXiv	S::LANLTopic:arXiv	HTTP	fetchId
BIOSIS	ERL	BX	S::ERL::BX	Z39.50	A
BIOSIS	ADVANCE	Biosis	S::ADVANCE::Biosis	Z39.50	fetchId
Wiley	Wiley	WIS	S::Wiley::WIS	none	none

Tavola 4: Alcune risorse SFX-aware con i loro *serviceDesc*, protocolli di cattura e chiavi di cattura

## Il componente SFX di servizio

Il compito del componente locale di servizio SFX comincia quando il meccanismo locale di reindirizzamento gli passa il contenitore di metadati, che contiene, in una rappresentazione coerente:

- i metadati della link-source che sono stati resi disponibili dal meccanismo locale di reindirizzamento
- l'informazione sull'origine della link-source, essenzialmente le informazioni del *serviceDesc*.

Il compito del componente di servizio SFX è inviare servizi estesi sulla base di queste informazioni: ecco alcune importanti considerazioni riguardo al componente di servizio SFX suggerite dall'esperimento Ghent&LANL:

- la quantità e qualità dei metadati della link-source che sono resi disponibili nel contenitore di metadati dipende dal tipo di risorsa da cui la link-source ha origine, e dalla quantità di informazione che l'authority che gestisce la risorsa permette o supporta che sia catturata. In certi casi tali metadati possono essere corrotti, o mancare di informazioni essenziali perché il processo valutativo SFX lavori bene;
- il componente di servizio SFX deve essere facilmente portabile tra diversi ambienti di biblioteca digitale e rimanere facilmente gestibile;
- il componente di servizio SFX deve in definitiva fornire i link di servizio in modalità just-in-time

Come si vede da una dettagliata descrizione del componente di servizio SFX, questi problemi sono stati affrontati per mezzo di:

- per a.: il GenericRequest object (l'oggetto di richiesta generico);
- per b.: una generalizzazione dell'implementazione del database SFX, che riflette esplicitamente la nozione di rilevanza globale e locale e di servizi concettuali, così come la nozione di soglie locali e globali;



- per c.: la soluzione del TargetParser (analizzatore del punto di arrivo).

## L'oggetto GenericRequest

Il componente di servizio prenderà in input il contenitore di metadati inviatogli dal meccanismo locale di reindirizzamento e lo trasformerà in una rappresentazione interna normalizzata, chiamata GenericRequest Object. La tavola 5 mostra una rappresentazione del GenericRequest Object per la terza citazione della figura 5. Il GenericRequest Object è un oggetto intelligente, che è in grado di autovalutare la validità dei suoi elementi informativi sulla base di regole predefinite. Esso può anche aumentare o migliorare il suo contenuto, utilizzando informazioni prese da un database di supporto. Per esempio, la citazione in figura 5 non contiene un numero ISSN né un titolo di rivista, ma piuttosto un titolo abbreviato. In questo caso il GenericRequest Object aumenta il suo contenuto aggiungendo le informazioni mancanti attraverso la comunicazione con un database di supporto. Ovviamente il GenericRequest Object contiene anche una versione normalizzata dei metadati della link-source, così come informazioni sulla sua origine.

Al momento dell'esperimento, l'interoperabilità tra il componente locale di servizio SFX ed eventuali meccanismi locali di reindirizzamento non-SFX, non era un problema, poiché non ne esisteva nessuno; pertanto, in Ghent&LANL, per ragioni di semplicità lo schema di metadati del GenericRequest Object ha assunto il ruolo di schema di metadati di interfacciamento tra il componente locale di reindirizzamento e il componente locale di servizio.

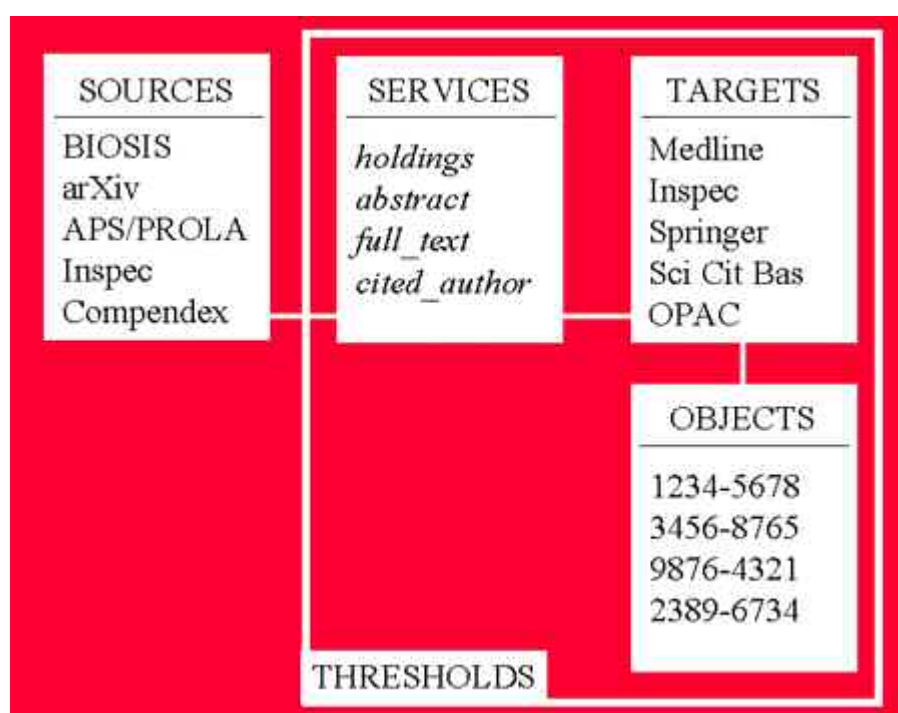
```
<perldata>
<hash>
<item key="rec$vendorId">Wiley</item>
<item key="rec$databaseId">WIS</item>
<item key="rec$dbId">Wiley::WIS</item>
<item key="objectType">JOURNAL</item>
<item key="@abbrevTitle">
<array>
<item key="0">N ENGL J MED</item>
</array>
</item>
<item key="journalTitle">NEW ENGLAND JOURNAL OF MEDICINE</item>
<item key="ISSN">0028-4793</item>
<item key="year">1994</item>
<item key="volume">330</item>
<item key="startPage">691</item>
<item key="endPage">7</item>
<item key="@authLast">
<array>
<item key="0">Saven</item>
<item key="1">Piro</item>
</array>
</item>
<item key="@authInit">
<array>
<item key="0">A</item>
<item key="1">L</item>
</array>
</item>
<item key="articleTitle">The newer purine analogues for the treatment of
hairy-cell leukemia.</item>
</hash></perldata>
```

**Tavola 5: Rappresentazione di un GenericRequest object aumentato per la link-source della [Figura 5](#)**

## Il servizio di linking SFX e la base SFX

Come risultato del processo sopra descritto, per il componente di servizio SFX è diventata disponibile una istanza del GenericRequest Object relativo alla link-source per cui sono stati richiesti servizi estesi. Il compito di questo componente sarà inviare i servizi estesi all'utente che li ha richiesti. In questo senso, il componente di servizio SFX è un servizio di linking che, dato un certo input "documento", dà in output altri documenti correlati a quell'input. Però è un servizio di linking speciale poiché non immagazzina relazioni statiche tra documenti individuali; piuttosto, immagazzina relazioni tra le risorse da cui originano i documenti. In SFX queste relazioni sono chiamate servizi concettuali e sono mantenute nella base SFX. Il processo di valutazione SFX determinerà la rilevanza di ognuno di questi servizi concettuali, sulla base delle informazioni e dell'origine della link-source.

Il requisito imposto all'implementazione del componente SFX in ambito Ghent&LANL, di essere facilmente portabile tra diverse biblioteche digitali, ha condotto ad un'importante



generalizzazione del progetto della base SFX, ottenuta riflettendo esplicitamente nell'implementazione la nozione di rilevanza globale e locale dei servizi. Una rappresentazione sintetica dell'organizzazione della base SFX di Ghent&LANL è mostrata in figura 7.

Figura 7: versione semplificata della SFX-base

### Dividere la tabella Colli

Come nella versione Elektron della base SFX, la tabella Source contiene le risorse informative che possono essere origini di link-source: si tratta di risorse SFX-aware.

La tabella Colli nella versione Elektron conteneva i servizi concettuali direttamente accoppiati con le risorse punto d'arrivo (target) (vedi [Tavola 2](#) in [Van de Sompel & Hochstenbach 1999b](#)): tale assetto non era sufficientemente generico, e nel progetto Ghent&LANL la Colli è stata divisa in due tabelle, una che mantiene il nome Colli, ed un'altra che è stata chiamata tabella dei target: quest'ultima contiene quelle risorse verso cui è possibile un link; la Colli, che connette la tabella delle risorse sources con la tabella delle risorse target, ora esprime solo il tipo di servizio che connette sources e target.

La tavola 6 mostra i tipi di servizi implementati in Ghent&LANL:

<u>Servizi Colli</u>	<u>Funzioni</u>
abstract	ricerca di un abstract in un database di indicizzazione e abstract per l'unità informativa (item) rappresentata
author	ricerca in un abs di riferimenti da un autore dell'item rappresentato dall'oggetto GenericRequest
cited_author	ricerca di citazioni bibl. a lavori di un autore menzionato nell'oggetto GenericRequest
cited_reference	ricerca di lavori che citano l'item rappresentato nell'oggetto GenericRequest
full_text	link al testo pieno dell'item rappresentato nell'oggetto GenericRequest
genoma	ricerca delle sequenze trovate nell'oggetto GenericRequest
holding	ricerca di informazioni di possesso in un opac per l'item rappresentato nell'oggetto GenericRequest
review	ricerca di una recensione a un libro per l'item rappresentato nell'oggetto GenericRequest

Tavola 6: Servizi del Colli e oro funzionalità

## Trarre vantaggio dalla rilevanza globale dei servizi concettuali

Non è una coincidenza che le risorse mostrate come source o come target portino i loro nomi globali comuni piuttosto che quello delle loro implementazioni locali a Ghent o a LANL: questo effettivamente riflette la considerazione che i servizi relativi alle risorse Source e Target hanno una rilevanza globale: è globalmente rilevante inviare un servizio “abstract” che, data una link-source da BIOSIS, mostri il corrispondente abstract di Medline. Tale servizio concettuale può essere immaginato indipendentemente dalle implementazioni di ognuna delle risorse in una specifica biblioteca digitale. Perciò la base SFX di Ghent&LANL esprime le relazioni tra Sources e Targets a un livello di rilevanza globale: esiste un servizio “abstract” che connette BIOSIS e Medline indipendentemente dalle loro implementazioni. Un limitato numero di esempi di come tali servizi a rilevanza globale connettano Source e Target è dato dalla tavola 7:

<u>COLLI</u>		
<u>Source (punto di partenza)</u>	<u>servizio del Colli</u>	<u>target (punto di arrivo)</u>
APS/PROLA	abstract	Inspe
the ArXiv	author	Inspe
BIOSIS	abstract	Medline
BIOSIS	genome	Genome Base
CC	abstract	LiSa
EconLit	review	Books in Print
Inspe	full_text	Springer
Wiley	abstract	Medline
Wiley	cited_reference	Science Cit. Base

Tavola 7: Esempi di relazioni di servizio tra Sources e Targets

## Localizzazione dei servizi di rilevanza globale

I servizi mostrati nella tavola 7, di rilevanza globale, non prendono in considerazione questioni di rilevanza in relazione alla specifica biblioteca digitale. La localizzazione (contestualizzazione a livello locale) dei servizi di rilevanza globale è ottenuta per mezzo di:

- l'introduzione, accanto ai nomi comuni globali, di campi riferiti all'implementazione locale: come mostrato nella tavola 8 e 9, viene aggiunta, accanto al nome globale comune della Source, una chiave che rispecchia il valore del *ServiceDesc* dell'implementazione locale della risorsa (trovato nel campo *rec\$dbID* dell'oggetto *GenericRequest*). Allo stesso modo, sul versante target, viene aggiunto, accanto al nome globale di cui il target locale è un'implementazione, il nome di un *TargetParser* (AnalizzatoreDiTarget) locale. La procedura *TargetParser* implementa una sintassi link-to nell'implementazione locale della risorsa target. Dal confronto tra le tavole 8 e 9 si nota che Ghent e LANL usano due diversi *SourceParser* per BIOSIS, il che rispecchia il

fatto che ne hanno due diverse implementazioni. E invece condividono il TargetParser per fornire il servizio “abstract” di Medline, poiché entrambe hanno scelto di fornirlo sull’implementazione PubMed.

- la disattivazione dei servizi di rilevanza globale quando non sono di rilevanza locale: quando il Source o il Target necessari per implementare un certo servizio non sono disponibili nella collezione digitale in esame, oppure quando l’implementazione locale del target non supporta il meccanismo di link-to necessario per implementarlo, o ancora, quando i bibliotecari decidono che quel servizio non è di utilità per i loro utenti finali, tale servizio viene disattivato (con un apposito flag): esso non verrà più preso in considerazione nel processo valutativo SFX che decide della rilevanza locale dei servizi concettuali. Nella tavola 8, questo è il caso dei servizi con Inspec come fonte, dato che Ghent non ha alcuna implementazione di Inspec nella sua collezione; nella tavola 9, è il caso dei servizi con LiSa come target, dato che LANL non ha accesso all’implementazione di LiSa.

SOURCE		COLLI	TARGET	
local	global		global	local
S::APS::PROLA	APS/PROLA	<i>abstract</i>	Inspec	T::ERL::IN
S::LANLTopic:arXiv	the arXiv	<i>author</i>	Inspec	T::ERL::IN
S::ERL::BX	BIOSIS	<i>abstract</i>	Medline	T::NCBI::PubMed
S::ERL::BX	BIOSIS	<i>genome</i>	Genome Base	T::NCBI::Genome
S::ERL::CCO	Current Contents	<i>abstract</i>	LiSa	T::ERL::LI
S::ERL::EC	EconLit	<i>review</i>	Books in Print	T::ERL::BOIP
inactive	Inspec	<i>full_text</i>	Springer	T::Springer::LINK
S::Wiley::WIS	Wiley	<i>abstract</i>	Medline	T::NCBI::PubMed
S::Wiley::WIS	Wiley	<i>cited_reference</i>	Science Cit. Base	T::CIC15:SciSearch

Tavola 8: Localizzazione dei servizi della [Tavola 7](#) a Ghent

Source		Colli	Target	
local	global		global	local
S::APS::PROLA	APS/PROLA	<i>abstract</i>	Inspec	T::ERL::IN
S::LANLTopic:arXiv	the arXiv	<i>author</i>	Inspec	T::ERL::IN
S::Advance::Biosis	BIOSIS	<i>abstract</i>	Medline	T::NCBI::PubMed
S::Advance::Biosis	BIOSIS	<i>genome</i>	Genome Base	T::NCBI::Genome
S::ERL::CCO	Current Contents	<i>abstract</i>	LiSa	inactive
inactive	EconLit	<i>review</i>	Books in Print	T::ERL::BOIP
S::Advance::Inspec	Inspec	<i>full_text</i>	Springer LINK	T::Springer::LINK
S::Wiley::WIS	Wiley	<i>abstract</i>	Medline	T::NCBI::PubMed
S::Wiley::WIS	Wiley	<i>cited_reference</i>	Science Cit. Base	T::CIC15:SciSearch

Tavola 9: Localizzazione dei servizi della [Tavola 7](#) a LANL

## Soglie globali e locali

La relazione tra risorse Source e Target, espressa da una connessione di servizio registrata nel Colli, è soggetta a restrizioni chiamate “soglie di attivazione”: esse servono a mettere a punto i servizi concettuali e minimizzare la presentazione di servizi inappropriati. Per illustrare il concetto, si descrivono due tipi di “soglie”:

1. “soglie” espresse in termini di vincoli sugli elementi di metadati che formano la struttura dell’oggetto GenericRequest: tecnicamente queste “soglie” sono espresse come proposizioni condizionali sui nomi di campi dell’oggetto GenericRequest: in molti casi sono semplicissime, ma possono essere script di qualsiasi livello di complessità, per esempio:
  - autore\_citato: perché tale servizio sia rilevante, la “soglia” minima che va oltrepassata è l’esistenza di un nome di autore nell’oggetto GenericRequest;
  - review: tale servizio è rilevante solo per una link-source che descriva un libro;

- genome: tale servizio è rilevante solo se l'oggetto GenericRequest contiene identificatori di sequenze di genomi;
- abstract: la "soglia" per un tale servizio dovrebbe esprimere che la link-source contenga un articolo e porti almeno informazioni di anno, volume, fascicolo:  
`$GenericRequestObject->need('objectType', 'eq', 'JOURNAL') &&`  
`$GenericRequestObject->need('year') && $GenericRequestObject->need('volume') &&`  
`$GenericRequestObject->need('issue')`

2. "Soglie di repertoriamento" (objectLookup): il servizio di abstract è chiaramente soggetto ad un altro tipo di limitazione, e cioè richiede che la risorsa target, verso cui si intende dirigere il link, indicizzi quel periodico in cui è stato pubblicato l'articolo riferito nell'oggetto GenericRequest: questo requisito spiega l'esistenza della tavola degli oggetti in figura 7 e di una speciale "soglia di repertoriamento"; questo tipo di "soglia" entra in campo anche per decidere della rilevanza di un servizio "full\_text" verso un dato archivio di testi pieni, determinando se tale archivio contiene il periodico che porta l'articolo citato nella link-source. Proprio come per i servizi concettuali, per questo tipo di "soglie" c'è un componente globale e un componente locale: la "soglia di repertoriamento" globale per un servizio "full\_text" verso la collezione di testi pieni della Springer considererà se un certo periodico è un periodico elettronico Springer o no; il componente locale di questa soglia considererà se il periodico è parte della collezione della biblioteca digitale. Nello stesso contesto, la soglia globale può esprimere il fatto che un periodico sia disponibile in formato elettronico dal 1996, mentre il componente locale può indicare che l'abbonamento locale parte solo dal 1998. Analogamente, il servizio di abstract da BIOSIS a Medline è soggetto ad una "soglia di repertoriamento" globale che esprime il fatto che l'anno di pubblicazione nell'oggetto GenericRequest deve essere posteriore al 1965, anno di partenza di Medline, mentre il componente locale può essere settato ad un anno più recente, se l'implementazione locale di Medline comprende un numero inferiore di dati.

## Il processo di valutazione SFX

Per presentare dei servizi estesi per un dato oggetto GenericRequest, il processo valutativo SFX determinerà la rilevanza di ciascuno dei servizi concettuali registrati nella base SFX utilizzando il contenuto, o la mancanza di esso, nell'oggetto GenericRequest.

FASE 1: selezione dei servizi attivi che abbiano come source la risorsa di origine espressa nell'oggetto GenericRequest.

L'interfaccia tra il componente di reindirizzamento e il componente di servizio invia sia i metadati della link-source che informazioni sull'origine della link-source; queste ultime sono registrate nel campo rec\$dbId dell'oggetto GenericRequest creato dal componente di servizio. In fase di valutazione, il valore di questo campo diventa la chiave per cercare nel componente locale della tabella Source della base SFX: lì viene reperito, accanto a questa chiave che si riferisce all'implementazione locale della risorsa, il suo nome comune globale, il quale viene poi connesso, attraverso i servizi del Colli, con diversi nomi globali di risorse Target, come mostrato nella tavola 7. Perciò, il risultato di questa ricerca è un insieme di servizi che potrebbero essere rilevanti per l'oggetto GenericRequest in esame, giudicando dalla sua origine. La disattivazione di alcuni servizi durante la localizzazione della base SFX garantisce che l'insieme risultante rifletta già la situazione della biblioteca digitale specifica.

Le tavole 8 e 9 mostrano in grassetto il meccanismo suddetto per un oggetto di GeneriRequest che rappresenta un item originatosi dall'implementazione locale di Ghent di BIOSIS: il suo valore di rec\$dbId è ERL::BX; in questa fase del processo di valutazione, S::ERL::BX (viene aggiunto un prefisso S come modo per riferirsi alla source) diventa la chiave per consultare il componente locale della tabella source: lì viene trovato il nome comune globale della risorsa: BIOSIS, a cui sono legati

vari servizi verso varie risorse Target, per esempio “abstract” connette BIOSIS con Medline, e “genome” connette BIOSIS con la Genome Base. Questi sono i servizi a rilevanza globale reperiti.

FASE 2: filtraggio dei servizi attivi selezionati, confrontando il contenuto dell’oggetto GenericRequest con le “soglie”.

La fase 1 del processo valutativo SFX esclude i servizi della base SFX che non hanno BIOSIS come origine. Per ognuno dei servizi che restano, l’informazione dell’oggetto GenericRequest sarà confrontata con le “soglie”, globali e locali, relative a quei servizi: il servizio “genome” che connette BIOSIS alla Genoma Base sarà escluso se l’oggetto GenericRequest non contiene dati per il parametro genID. Il servizio “abstract” che connette BIOSIS con Medline sarà escluso se la “soglia di repertoriamento” sul valore ISSN dell’oggetto GenericRequest indica che il periodico in questione non è indicizzato in Medline, o anche se l’oggetto GenericRequest non contiene valori per anno, volume e fascicolo. Di nuovo, alcune “soglie” esprimono situazioni locali, e poiché esse possono sovrascrivere quelle globali, il risultato di questa operazione di filtro rifletterà la situazione della specifica biblioteca digitale. I servizi selezionati dalla fase 1, per i quali fallisca anche solo una delle valutazioni di “soglia”, vengono esclusi come non rilevanti. Quelli che passano attraverso tutto il processo di valutazione, saranno presentati all’utente nello schermata di menù SFX come servizi estesi rilevanti per l’oggetto GenericRequest in esame, quindi per la link-source da cui tutto il processo ha preso inizio al momento del click sul bottone SFX.

## **Risolvere, attraverso TargetParser, i servizi estesi a rilevanza locale in URL**

Coerentemente con la filosofia di linking just-in-time di SFX, l’insieme di servizi rilevanti ottenuto come risultato del processo di valutazione SFX sopra descritto, non sono risolti in URL al momento della loro presentazione all’utente che ha lanciato la richiesta di servizi estesi; piuttosto, per ogni voce del menù della schermata di menù SFX vengono inviati i seguenti elementi, come parametri per uno script che partirà al momento in cui l’utente selezionerà quella voce di menù:

- l’identificatore dell’oggetto GenericRequest;
- un nome che si riferisce al servizio rappresentato dalla voce di menù e al suo target;
- per alcuni servizi, elementi di metadati sovrascrivibili provenienti dall’oggetto GenericRequest (vedi le schermate di menù SFX nella [Figura 8](#), [Figura 10](#), [Figura 12](#) e [Figura 14](#)).

Quando l’utente clicca su una voce di menù, viene lanciato l’appropriato script TargetParser che corrisponde al servizio e al target scelto. Questi TargetParser implementano sintassi link-to specifiche per ogni risorsa: essi prendono in input i dati dall’oggetto GenericRequest e calcolano l’URL a cui l’utente può essere reindirizzato.

## **Commenti**

### **L’impatto del ridisegno del componente di servizio SFX**

Il nuovo disegno del componente di servizio SFX che riflette aspetti di rilevanza globale e locale, ha un impatto considerevole sulla portabilità e maneggevolezza del componente di servizio SFX. Una volta che è stata compilata una base SFX contenente servizi concettuali di rilevanza globale, con appropriate soglie globali, la localizzazione dell’assetto richiede una fatica minima.

A illustrazione di questo, è interessante considerare ancora una volta l’esempio di servizio abstract da BIOSIS verso Medline: questo servizio è stato inizialmente localizzato a Ghent inserendo i nomi del SourceParser locale per BIOSIS e del TargetParser locale per Medline. I due parser implementano la connessione desiderata con la piattaforma locale ERL della SilverPlatter, che ospita i database locali. LANL ha una implementazione diversa di BIOSIS, che attualmente gira su un sistema Geac, Advance, e non ha alcuna implementazione locale di Medline: perciò ha scelto



di usare come target l'implementazione libera di Medline, PubMed. Quando si è trasportata a LANL la base SFX che era stata inizialmente localizzata a Ghent, sono state sufficienti, per attivare il servizio nel nuovo ambiente, operazioni di editing molto leggere sulla base SFX: i servizi globali e le soglie globali restano validi; i valori dei parser specifici per la versione di BIOSIS e di Medline di Ghent sono stati sovrascritti con quelli specifici per Los Alamos, come si vede in tavola 9. La soglia che indicava che Medline era disponibile a Ghent solo al 1985 è stata sovrascritta per LANL dalla soglia 1965 (rendendola così uguale alla soglia globale). L'eleganza del meccanismo di link-to di PubMed, *Entrez*, e la disponibilità sulla versione libera dell'intera collezione Medline, ha indotto Ghent a riconsiderare il target da usare (e quindi il TargetParser) a favore dell'implementazione PubMed: a questa decisione, di nuovo, ha corrisposto una correzione molto limitata nella base SFX di Ghent.

In Ghent&LANL la maggior parte dei TargetParser sono implementati come Script in Perl. Verso la fine dell'esperimento, si è approfittato della possibilità di lanciare preliminarmente una versione del calcolatore (programma di calcolo) S-Link-S ([Openly Inc. 1999](#)), il quale è progettato per calcolare URL sulla base di metadati e template XML che descrivono la sintassi link-to in un modo compatibile con S-Link-S ([Hellman 1998](#)); utilizzando questo strumento, gli script TargetParser SFX che calcolano la URL possono essere sostituiti da template XML che descrivono la sintassi link-to e vengano usati come input per il calcolatore S-Link-S. L'esperimento si è concluso con una soluzione ibrida, in cui il componente di servizio SFX, per calcolare la URL, è capace di scegliere dinamicamente tra i due meccanismi, entrambi disponibili: il TargetParser o il template + calcolatore S-Link-S. Il TargetParser può essere condiviso tra le diverse biblioteche digitali che hanno bisogno di link verso le stesse risorse o famiglie di risorse: anche questo riduce gli investimenti per installare la soluzione SFX. Se poi si affermasse la connessione tra SFX e S-Link-S, la gestione della soluzione SFX sarebbe ancora più leggera, dato che sarebbero gli editori a fornire i templates link-to e i metadati corrispondenti allo schema S-Link-S. La condivisione del TargetParser sarebbe allora sostituita dall'uso dei Template S-Link-S secondo lo schema che è già stato fissato da Eric Hellman.

Anche i SourceParser sono facilmente trasferibili, con poche modifiche, tra diverse biblioteche digitali. Per esempio, i sistemi Opac in tutto il mondo supportano il protocollo Z39.50 e rispondono alle ricerche con record in formato Marc. Facendo astrazione dalle colpevoli idiosincrasie tra le varie implementazioni di Z39.50 e di Marc, il SourceParser per un opac di questo tipo può essere riutilizzato con pochi cambiamenti per ogni situazione locale, a parte gli adattamenti dei parametri Z39.50 quali il server, la porta, il client. Ancora, lo Z39.50 può essere usato per catturare la link-source da tutti i database sulla piattaforma ERL e anche qui cambieranno solo i parametri Z39.50. Tutte le implementazioni di MathSci sulla piattaforma ERL possono usare le stesse procedure di analisi, rendendo la parte di analisi del SourceParser quasi universale. Questo è anche il caso del SourceParser usato per l'archivio APS/PROLA e per i periodici Wiley, poiché esiste un'unica implementazione di essi. Questo approccio apre attraenti prospettive di condivisione su larga scala dei SourceParser, riducendo ulteriormente gli investimenti necessari per installare la soluzione SFX. Inoltre esso permette ai fornitori di informazione di fornire SourceParser per le loro risorse, mantenendo in tal modo il controllo sull'informazione che viene catturata.

## **L'impatto di un componente di servizio basato su servizi concettuali**

Nell'esperimento Ghent&LANL, proprio per la sua natura complessa e distribuita, hanno assunto piena evidenza le conseguenze dell'introduzione, avventura nell'esperimento Elektron, di un servizio di linking basato su un database di servizi concettuali. Attualmente, più risorse vengono aggiunte all'ambiente, più si rivela l'eleganza e la efficacia della soluzione: l'introduzione nell'ambiente di una nuova risorsa SFX-aware richiede un editing molto limitato della base SFX

perché tutti i servizi concettuali esistenti, già registrati nella base SFX, diventino immediatamente disponibili anche per la nuova risorsa. E' semplicemente notevole la maniera dinamica in cui SFX presenta una lista di servizi estesi per la link-source della risorsa appena aggiunta. Anche i progettisti del sistema trovano sempre più difficile predire il risultato di una richiesta di servizi estesi, pur conoscendo il sistema e i database, in entrata e in uscita, su cui poggia, e anche quando studiano in dettaglio il contenuto della link-source o del suo oggetto GenericRequest. A illustrazione di questo, basti menzionare il fatto che SFX invia un link verso PubMed da una citazione nell'Information Science Journal JASIS, per cui lo stesso Wiley non ha alcun link: Wiley inserisce link statici verso PubMed, per le citazioni nei suoi periodici; ma probabilmente lo fa solo per quelli di soggetto biomedico, poiché inviare tutte le citazioni al processo NCBI OubRef sarebbe svantaggioso a livello di costi/benefici. L'approccio SFX, dinamico e concettuale, non richiede tale processo di precomputazione ed è in grado di riconoscere al volo l'opportunità di presentare un link verso PubMed per una citazione da JASIS. Un altro esempio notevole ed invitante è quando i link al full-text presentati da SFX conducono da una citazione dentro un periodico Wiley ad un articolo in un altro periodico Wiley: al momento dell'esperimento, Wiley non offriva tale servizio di collegamento all'interno della propria collezione, di cui pure aveva il controllo completo. Questi esempi non solo illustrano la potenza della soluzione SFX ma, più importante, danno una decisiva indicazione dei problemi di scala delle soluzioni a link statici.

## **Il meccanismo di reindirizzamento SFX e gli identificatori specifici di namespace**

Sono in corso sforzi significativi per rendere possibile il collegamento da citazione tramite DOI ([Paskin 1999a](#)). Gli editori forniranno al metaDatabase DOI i metadati delle loro pubblicazioni insieme al codice DOI corrispondente; altri editori potranno perciò confrontare le citazioni nei loro articoli con il metaDatabase DOI e inserire nella loro citazione il corrispondente codice DOI. Attualmente, poiché i più diffusi browser non supportano il protocollo [handle](#), tale codice DOI viene linkato come `<http://dx.doi.org/<codice>>`, e il proxy di gestione DOI risolverà questo link in una URL univoca che è quella della pubblicazione al sito dell'editore ([Paskin 1999b](#)). Questo link è un perfetto esempio di link chiuso che non tiene conto del contesto locale in cui viene utilizzato ([Van de Sompel & Hochstenbach 1999a](#)): questo tipo di link ha dato origine al problema di Harvard, poiché non ha tenuto conto della possibilità che gli stessi documenti possano venire archiviati in un altro sito, preferito; ancora, questo meccanismo non consente di fornire per una certa citazione altri servizi estesi, localmente rilevanti, poiché la loro fornitura richiede tutti i metadati e non solo l'identificatore della citazione.

L'approccio di reindirizzamento locale presentato nel lavoro SFX prospetta un modo pragmatico di aprire uno schema di link chiusi: DOI può essere incluso in una URL SFX che punta al server SFX preferito: p.e., per una citazione in un articolo Wiley InterScience, che ha un link statico DOI "10.1000/123456789", il link può essere riscritto dinamicamente se si rileva l'esistenza di un componente di servizio. Nel caso di un componente di servizio SFX, può diventare:

[http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorID=Wiley&databaseId=WIS&nameSpace=DOI&objectDesc=Urlencode\(DOI=10.1000/123456789\)](http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorID=Wiley&databaseId=WIS&nameSpace=DOI&objectDesc=Urlencode(DOI=10.1000/123456789))

In sostanza, tale meccanismo pragmatico può reindirizzare l'identificatore al componente di reindirizzamento di un sistema di risoluzione selettiva, che può decidere che cosa fare di esso sulla base del contesto locale. Nel caso di SFX, il ricevimento della URL suddetta lancia il SourceParser: come visto sopra, sarà il SourceParser corrispondente alla parte serviceDesc della URL; ancora, al momento del rilevamento del parametro nameSpace, questo parametro di default può essere sovrascritto e diventare il SourceParser specifico per quel nameSpace (in questo esempio il sourceparser per DOI). Questo SourceParser per il DOI farà una cosiddetta ricerca inversa nel metaDatabase DOI, usando il valore del DOI come chiave per catturare i metadati corrispondenti. Sia tali metadati che il DOI possono allora essere usati nel processo che determina i servizi estesi

rilevanti per la data citazione, incluso il link al full-text più appropriato. Come sarà discusso nella sezione successiva, per altri tipi di componenti di servizio, sarebbe sufficiente il reindirizzamento del DOI senza bisogno di accedere ai metadati.

Lo stesso meccanismo può essere usato per aprire link connessi a citazioni che portano identificatori di altri domini (Namespace), quali PubMed o Astrophysics Data System. Questo si vede facilmente prendendo la citazione seguente da un periodico Wiley:

Rainer RO, Geisinger KR. Beyond sensitivity and specificity. Am J Clin Pathol 1995; 103: 541-2. [Medline](#)

che ha un link statico a PubMed, che punta a:

<http://www4.ncbi.nlm.nih.gov:80/htbin-post/Entrez/query?uid=95259660&form=6&db=m&Dopt=r>

ed utilizza l'identificatore PubMed come chiave di ricerca in NCBI PubMed; esso può essere diretto verso un risolutore locale, se viene dinamicamente riscritto come:

<http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?>

[vendorID=Wiley&databaseId=WIS&nameSpace=Medline&objectDesc=URLencode\(Medline=95259660\)](http://isiserv.rug.ac.be/cgi-bin/sfx/bin/menu.cgi?vendorID=Wiley&databaseId=WIS&nameSpace=Medline&objectDesc=URLencode(Medline=95259660))

In maniera analoga all'esempio DOI, questa URL può inviare l'identificatore PubMed a un componente locale di reindirizzamento: nel caso di SFX, questo causerebbe il lancio del SourceParser specifico per il namespace Medline, che catturerebbe il record corrispondente dal database Medline; questo SourceParser può attualmente catturare i metadati sia dalla implementazione PubMed (usando il protocollo HTTP e la sintassi link-to *Entrez*) sia dall'implementazione locale di Medline, se ne esiste una.

Una volta che i metadati sono stati trasformati nell'oggetto GenericRequest, il processo di valutazione SFX può inviare i servizi estesi localmente rilevanti. In questo modo, il servizio iniziale Wiley è aumentato con servizi estesi localmente rilevanti. Ancora, la qualità dei metadati risultante dopo una tale cattura sarà migliore che non la citazione originaria, come si vede dal fatto che in essa non c'è il numero di fascicolo, (fatto comune nella letteratura medica) mentre il record catturato lo contiene insieme ad altra informazione di valore.

Queste considerazioni illustrano come i metadati della link-source non debbano necessariamente essere presi dalla risorsa d'origine: in entrambi gli esempi, per una link-source originatasi da un periodico Wiley, i metadati sono stati presi dalla risorsa autorevole per il dominio (namespace) di cui la link-source porta l'identificatore.

## Identificatori, metadati e componenti

E' interessante riflettere ulteriormente sulla natura del componente di servizio e sui requisiti che esso impone al meccanismo di reindirizzamento di una soluzione di risoluzione selettiva.

Per cominciare, prendiamo in considerazione i componenti di servizio che mirano unicamente all'invio della copia appropriata di full-text per una data link-source: un tale componente di servizio può operare semplicemente sulla base di un database di identificatori, il che configura un servizio di linking tradizionale, con link statici tra i documenti; per un tale componente di servizio, è sufficiente che il meccanismo di reindirizzamento trasferisca identificatori senza perdere tempo coi metadati associati; e poiché l'unico problema che deve affrontare è quello di Harvard, è sufficiente che esso contenga un archivio di identificatori più indirizzi dei testi pieni per cui è disponibile un accesso preferito diverso da quello di default. E però, tale archivio di identificatori è soggetto a diventare presto molto grande e difficile da mantenere, ed inoltre, nel caso in cui non ci sia alcun archivio locale di full-text ma solo un accesso preferito tramite un aggregatore esterno, può essere considerato assurdo per un'istituzione doverlo mantenere.

Queste considerazioni conducono alla desiderabilità di un componente di servizio di natura più astratta (generale), che sia costruito sulla logica implicita nella distribuzione delle collezioni

piuttosto che su identificatori individuali dei materiali nella collezione. In condizioni normali, tale logica può stabilire che tutti i periodici di un certo editore sono accessibili in un certo archivio elettronico, che certi ISSN devono essere acceduti in un altro archivio, e che per un certo numero ISSN deve essere consultato un archivio fino ad una certa data e un altro dopo tale data.

Questo livello di astrazione riduce drasticamente l'informazione che va mantenuta nel componente di servizio e perciò lo rende più scalabile (modulare). ma richiede che i metadati della link-source agiscano come operatori e non solo come identificatori. Se si aggiunge che, per link-source di origine varia (database bibliografici ma anche archivi di testi pieni, Opac ma anche archivi di e.print) gli identificatori necessari per produrre lo scenario prospettato non sono affatto disponibili e non lo diventeranno in tempi brevi, bisogna concludere che il componente di servizio dovrà essere in grado di operare con i metadati, considerando gli identificatori un particolare caso di metadato. Questo impone al componente di reindirizzamento il requisito di essere capace di inviare i metadati, non solo gli identificatori, della link-source.

Aggiungendo al compito del componente di servizio l'invio di altri servizi estesi, è difficile immaginare una soluzione modulare in un ambiente altamente distribuito che venga costruita su un'architettura con un database di link statici: l'esperimento Ghent&LANL ne ha dato parecchie prove; è perciò necessario un componente di servizio più astratto e dinamico, che sappia prendere decisioni sulla base di alcune regole tese alla valutazione della rilevanza di servizi concettuali, come illustrato nel lavoro SFX. Come risulta chiaro dall'esperimento SFX, questo tipo di componente di servizio richiede, per funzionare, la disponibilità dei metadati della link-source, e di nuovo questo impone al meccanismo di reindirizzamento di saper passare i metadati della link-source. Questo non significa che gli identificatori siano irrilevanti per questo tipo di soluzione: al contrario, è stato dimostrato che gli identificatori, provenienti da qualsiasi namespace, sono strumenti utili perché il componente locale di reindirizzamento possa passare dei metadati di alta qualità.

La conclusione generale di quanto sopra è che, realisticamente, gli identificatori non saranno sufficienti per affrontare il problema di fornire servizi estesi in un ambiente distribuito di biblioteca digitale: perché funzionino componenti di servizio modularizzabili sono necessari i metadati.

Muovendo da sinistra a destra nella scala dei componenti di servizio che vanno dai più tradizionali sistemi di linking statico ai sistemi di linking dinamico che si costruiscono su servizi concettuali, i dati richiesti perché i componenti di servizio svolgano le loro funzioni adeguatamente, vanno dagli identificatori ai metadati completi.

## **Illustrazione dei risultati del progetto**

I risultati concreti del progetto vengono mostrati in un'animazione di Lotus Screencam che mostra come un utente di Ghent e uno di LANL navigano nelle loro rispettive collezioni digitali: sono eseguibili stand-alone e girano solo su computer WinTel. Sono file grandi, perciò ne vengono dichiarate le dimensioni, e non contengono audio. Oltre a questi, vengono dati anche alcuni esempi con screendump. [Per tutto questo si rimanda all'[originale](#), N.d.T.].

]

## **Conclusioni**

L'esperimento [SFX@Ghent](#) & SFX@LANL ha condotto ad importanti generalizzazioni dei due componenti, già introdotti nell'esperimento Elektron, che sono essenziali per sistemi che debbano supportare la risoluzione selettiva: il meccanismo di reindirizzamento e il componente di servizio. Sebbene siano stati discussi in relazione reciproca, si è anche mostrato che essi possano essere componenti separati che scambiano informazioni in un formato unico di scambio di metadati.

Per l'esperimento, questo formato è stato definito internamente e ispirato alla struttura dell'oggetto GenericRequest poiché, in assenza di componenti locali di reindirizzamento non-SFX, l'interoperabilità a questo livello non era richiesta: se si rendessero disponibili diverse soluzioni locali di reindirizzamento e diverse soluzioni locali di servizio, la standardizzazione di questo

formato di scambio diventerebbe ineludibile. Sarebbe necessaria anche la standardizzazione della URL locale di reindirizzamento (la SFX-URL), nonché dei valori: ID del fornitore, ID del database e Namespace: una tale standardizzazione avrebbe impatto positivo anche su altri lavori in corso nell'ambiente della biblioteca digitale.

In sostanza, il meccanismo di reindirizzamento SFX può essere combinato con componenti di servizio ad architettura molto diversa, anche costruiti su un database di identificatori a linking statici. L'implementazione corrente del meccanismo locale di reindirizzamento SFX è costruito sul meccanismo CookiePusher, su una coerente URL SFX e sul SourceParser. Ognuno di questi blocchi costruttivi può essere sostituito da alternative più robuste, purché svolgano la stessa funzione. Gli investimenti richiesti per rendere un sistema SFX-aware usando il CookiePusher e la URL SFX sono stati minimizzati, tuttavia, sarà più facile implementare la conformità a SFX in risorse che inviino l'informazione in modo dinamico che non statico. E' stato mostrato che il meccanismo locale di reindirizzamento SFX può essere utilizzato per reindirizzare identificatori di specifici schemi (namespaces) a un componente di servizio locale; questo indica la capacità della URL SFX di aprire schemi di linking chiusi, il che è una potente dimostrazione della praticabilità dell'approccio.

Il componente di servizio SFX può operare anche con un metodo di reindirizzamento diverso, purché esso supporti l'invio dei metadati della link-source e della sua risorsa d'origine al componente di servizio. L'ambiente informativo Ghent&LANL, con le sue diverse risorse e le diverse tecnologie su cui esse girano, ha condotto ad un progetto in cui il servizio di linking SFX è diventato un modulo del tutto neutrale (autonomo) della biblioteca digitale, il quale può potenzialmente interoperare con ogni altro sistema dell'ambiente. La sua riprogettazione, che riflette la nozione di rilevanza globale e locale dei servizi, ha condotto ad una significativa riduzione degli investimenti necessari per applicare la soluzione; il carico di amministrazione del sistema è ulteriormente ridotto dalla possibilità di condividere SouceParser, TargetParser e template S-Link-S.

Per quanto si sa, Ghent&LANL è stato il primo esperimento che ha sviluppato un linking bi-direzionale context-sensitive tra risorse distribuite e gestite da diverse authorities, nell'ambiente dell'informazione accademica. Come si vede dagli esempi, rendere i sistemi SFX-aware e fornire, per link-source che originano da quei sistemi, servizi estesi attraverso la schermata di menù SFX, produce un ambiente informativo completamente ipertestuale in cui diventa possibile navigare tra risorse correlate distribuite, auspicabilmente nel modo che Gardner aveva immaginato ([Gardner 1990](#)): come per il più rinomato sistema ipertestuale, lo World Wide Web, la facilità di navigazione può indurre a perdersi nello spazio informativo, ma questa caratteristica è da considerarsi un pregio per la nostra soluzione, dato che una capacità di navigazione paragonabile non era mai stata sviluppata prima.

## **Direzioni future**

La soluzione SFX entrerà in produzione sia a Ghent che a Los Alamos verso il gennaio 2000. Questo auspicabilmente produrrà feedback dagli utenti, il che finora è stato limitato e comunque non investigato in maniera sistematica; produrrà anche un maggior coinvolgimento dei bibliotecari, il che non può che condurre a migliorare le aspettative degli utenti verso il sistema. La corrente soluzione SFX è pronta anche per essere testata (beta-testing, cioè test di secondo grado) in biblioteche digitali che siano gestite da staff con una ragionevole competenza tecnica; e però, per poter coordinare al meglio tale fase di test sarebbe auspicabile la presenza di qualche risorsa a livello centrale.

SFX viene intanto testato anche come strumento per integrare i meccanismi della comunicazione accademica "sovversiva" con quella convenzionale: questo, tra l'altro, viene fatto nel lavoro prototipo UPS, che mira a costruire un servizio di digital library multidisciplinare per le maggiori iniziative di e-print: in tale progetto, il lavoro SFX viene combinato con la ricerca Smart



Object Dumb Archive; è stato intrapresa anche un'esplorazione sulla possibilità di rendere SFX-aware il sistema SLAC/SPIRES HEP ([Harnad 1999](#)); infine, il progetto JISC/NS, che investiga il linking tra e all'interno delle citazioni negli e-print arXiv, intende sperimentare l'SFX come blocco integrato.

Su un altro versante, sono in corso discussioni con gli editori scientifici a proposito della interoperabilità con SFX: si ritiene la prospettiva di una più vasta distribuzione della soluzione del tutto realistica, ma probabilmente richiederà un supporto commerciale o in alternativa investimenti finanziari.

Il disegno corrente della base SFX in Ghent&LANL, implementando la nozione di rilevanza locale e globale, suggerisce anche la possibilità di una ristrutturazione architetturale che operi la divisione tra un componente centrale che descrive il livello globale, e un componente locale per la localizzazione: un tale ridisegno sposterebbe la soluzione SFX nella categoria 2 della categorizzazione in tavola 1 dei sistemi che supportano la risoluzione selettiva, e ridurrebbe notevolmente la ridondanza informativa nelle diverse implementazioni delle basi SFX, ulteriormente semplificando la gestione locale.

La ricerca sta per produrre un sistema di raccomandazioni basate sulle attività dell'utente in ambiente SFX: il fatto che possano essere seguiti ampi percorsi degli utenti attraverso risorse sia interne che esterne, arricchisce i dati di registrazione che possono essere sfruttati nel sistema di raccomandazioni. Infine, è necessaria una ricerca nell'area dell'inserimento dei bottoni SFX in contenitori di informazione chiusi quali i file PDF o Word.

## Riferimenti bibliografici

- Caplan, Priscilla. 1999a. A model for reference linking. Report of the working group of the reference linking workshop; May 1999. [<http://www.lib.uchicago.edu/Annex/pcaplan/reflink.html>].
- Caplan, Priscilla. 1999b. Report of the second workshop on linkage from citations to journal literature; June 9th 1999, Boston. [<http://www.niso.org/linkrept.html>].
- Caplan, Priscilla and William Y. Arms. 1999. Reference linking for journal articles. *D-Lib Magazine* 5, no. 7/8. [<http://www.dlib.org/dlib/july99/caplan/07caplan.html>].
- Gardner, William. 1990. The electronic archive: scientific publishing for the 1990s. *Psychological Science* 1, no. 6.
- Halstead, Amy. 1999. PROLA: More Than Just a Pretty Acronym. *APS News* 8, no. 8. [<http://www.aps.org/apsnews/0899/089914.html>].
- Harnad, Stevan. 1999. Integrating and navigating eprint archives through citation-linking (NSF / JISC-eLib Collaborative Project). [<http://www.princeton.edu/~harnad/citation.html>].
- Hellman, Eric. 1998. Scholarly Link Specification Framework (SLinkS). [<http://www.openly.com/SLinkS/>].
- Maly, Kurt, Michael Nelson, and Mohammad Zubair. 1999. Smart Objects, Dumb Archives: A User-Centric, Layered Digital Library Framework. *D-Lib Magazine* 5, no. 3. [<http://www.dlib.org/dlib/march99/maly/03maly.html>].
- Needleman, Mark. 1999. Meeting report of the NISO linking workshop; February 11th 1999, Washington DC. [<http://www.niso.org/linkrpt.html>].
- Okerson, Ann and James O'Donnell. 1995. Scholarly Journals at the Crossroads; A Subversive Proposal for Electronic Publishing. Washington, DC: Association of Research Libraries. [<http://www.arl.org/scomm/subversive/toc.html>].
- Openly Inc. 1999. S-Link-S Calculator. June 1999. [<http://www.openly.com/SLinkS/Calculator/>].
- Paskin, Norman. 1999a. DOIs used for reference linking. Washington & Geneva. [<http://dx.doi.org/10.1000/143>].
- Paskin, Norman. 1999b. DOI: Current Status and Outlook. *D-Lib Magazine* 5, no. 5. [<http://www.dlib.org/dlib/may99/05paskin.html>].
- Shishir, Gunavaram. 1996. CGI Programming on the World Wide Web. Sebastopol, CA.: O'Reilly and Associates, Inc.
- Spilka, Susan. 1999. Wiley InterScience Update. June 1999. [<http://www.wiley.com/about/corpnews/wisupdate.html>].
- Van de Sompel, Herbert. 1999. the Universal Preprint Service initiative. July 1999. [<http://vole.lanl.gov/ups/>].
- Van de Sompel, Herbert and Patrick Hochstenbach. 1999a. Reference linking in a hybrid library environment. Part 1: Frameworks for linking. *D-Lib Magazine* 5, no. 4. [[http://www.dlib.org/dlib/april99/van\\_de\\_sompel/04van\\_de\\_sompel-pt1.html](http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html)].
- Van de Sompel, Herbert and Patrick Hochstenbach. 1999b. Reference linking in a hybrid library environment. Part 2: SFX, a generic linking solution. *D-Lib Magazine* 5, no. 4. [[http://www.dlib.org/dlib/april99/van\\_de\\_sompel/04van\\_de\\_sompel-pt2.html](http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html)].